

## AN INFORMATION MANAGEMENT SYSTEM FOR DNA SEQUENCING

Arthur Kobayashi ([kobayashi1@llnl.gov](mailto:kobayashi1@llnl.gov)), David J. Ow, T. Mimi Yeh, Mark C. Wagner, Thomas R. Slezak. Human Genome Center, Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, Livermore, CA 94550.

We are implementing a comprehensive system to track and manage DNA sequencing data. The major goals of this system are to manage a complete source, processing, and analysis history for DNA sequenced in our local laboratory, and to streamline the flow of information and products through the laboratory.

To accomplish these goals we have developed tools to define and build clone library hierarchies to track detailed information at each processing and sequencing step, and other tools to help setup and archive results from our laboratory instruments. All information is stored in a relational database (Sybase). We have designed the system to be compatible with existing laboratory functions and protocols, minimize data entry, provide for consistent naming conventions, automate generation of sample sheets and setup files, and to track replications (reprocessing) of items.

There are numerous components that make up this system. The main interface, which runs on a Unix workstation, is used to create and maintain processing and sequencing information. Macintosh-based programs edit labelling or sequencing runs, update the database, and create final sample sheets or configuration files used to set up labellers and sequencers. We also use custom forms on Web browsers (such as Mosaic or Netscape) to implement specialized functions such as creating clone library entries or editing certain types of sequencing run assignments. We have developed other tools to streamline ABI setups and data archival using various Macintosh scripting tools.

We have been implementing this system using a variety of methods, including C, X Windows/Motif, UIMX (a commercial graphical-user interface builder), custom Excel spreadsheets, Sybperl, MacPerl, HTML (Hypertext Markup Language), and AppleScript. The resulting system is distributed, heterogeneous, and very loosely-coupled. Functions may be implemented as stand-alone, single-purpose programs, or as larger, multi-function programs; all of the pieces of the system are integrated over our computer network and share the use of an underlying common schema and relational database. This approach gives us great flexibility in selecting the most effective means available to implement a given function, and to replace or modify portions of the system without impacting other parts of the system.

The system is now in routine use in our laboratory. We are continuing to refine and improve existing functions, and are currently adding capabilities to streamline and integrate the analysis of sequenced DNA.

(This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-ENG-48.)